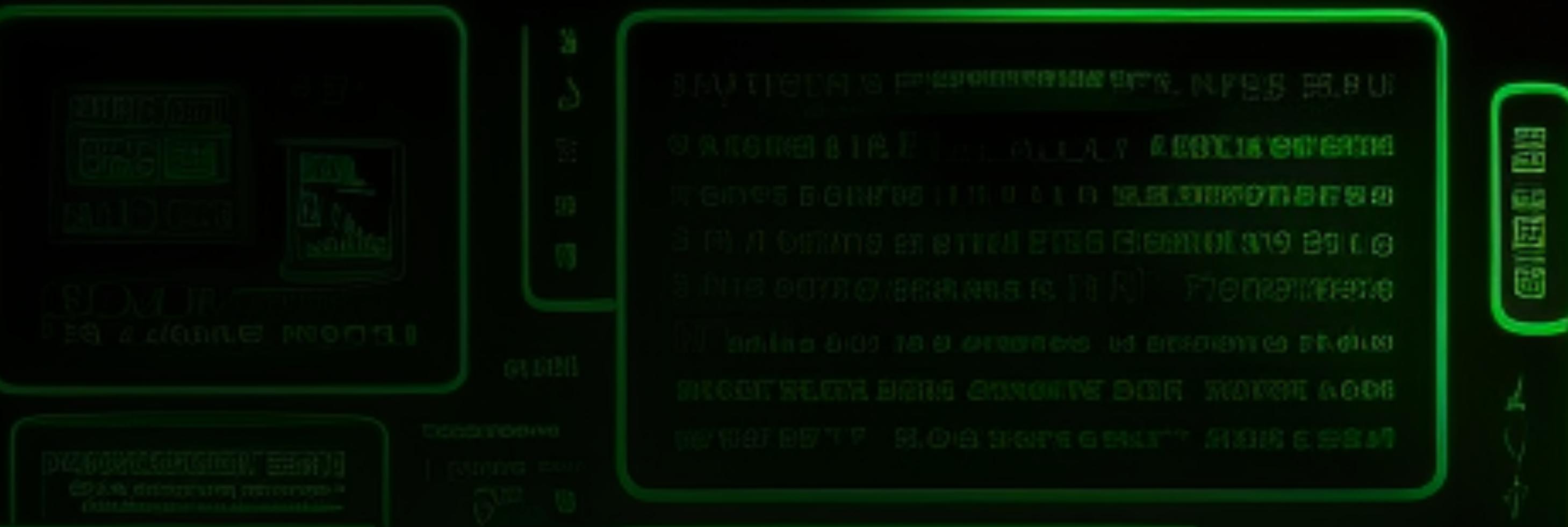


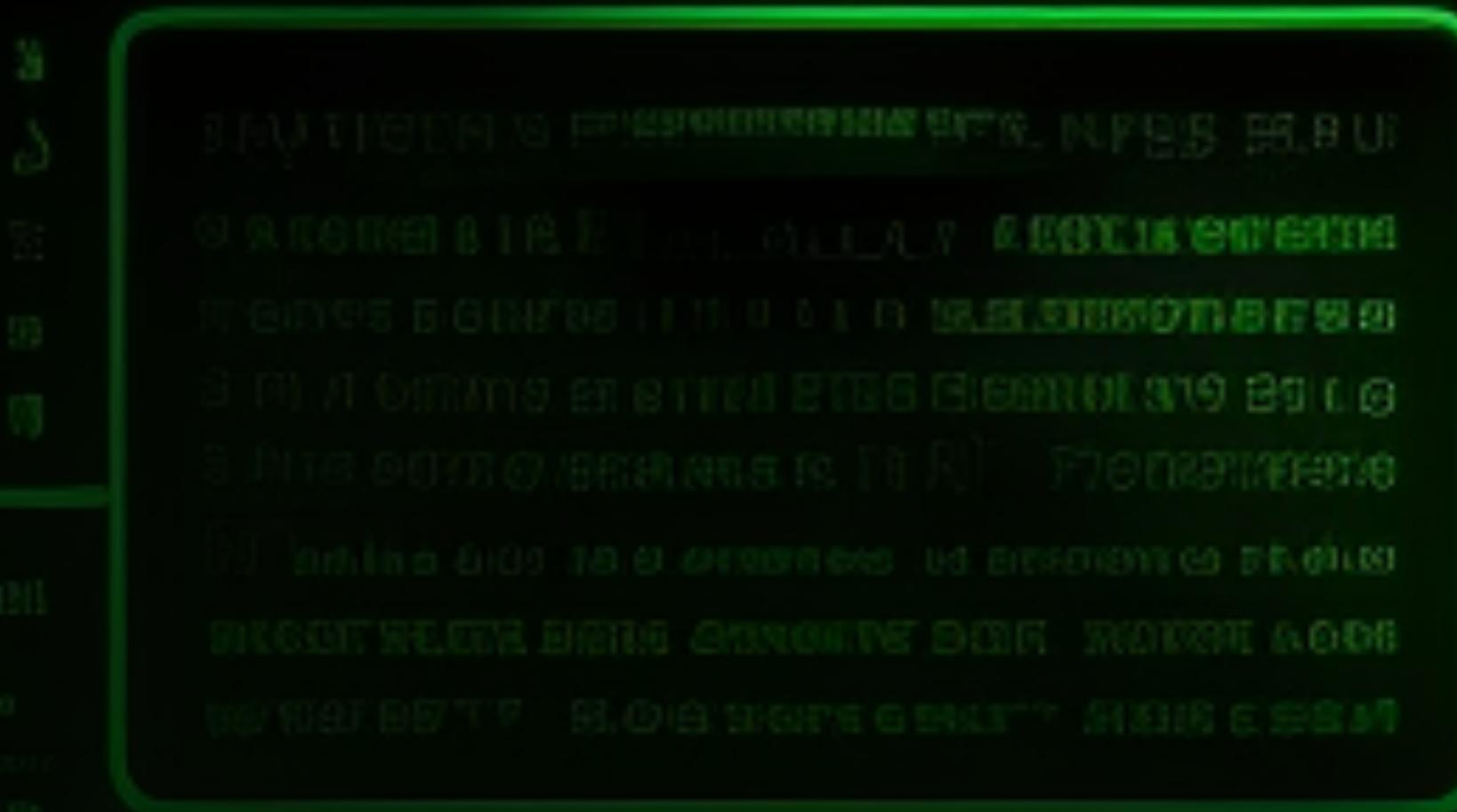
# Computational Social Science

Introduction

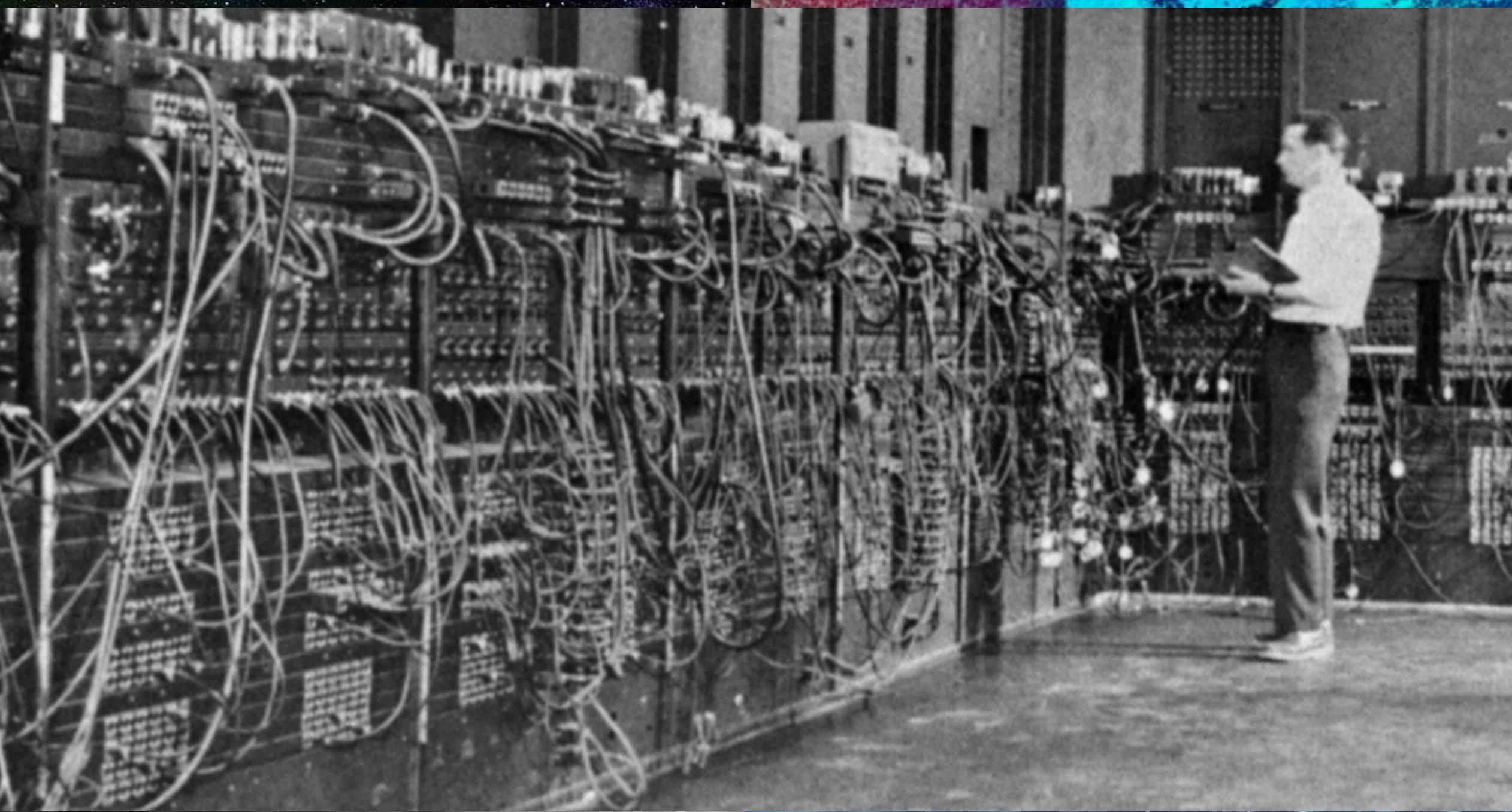
# What is computational social science?



# What is computational science?

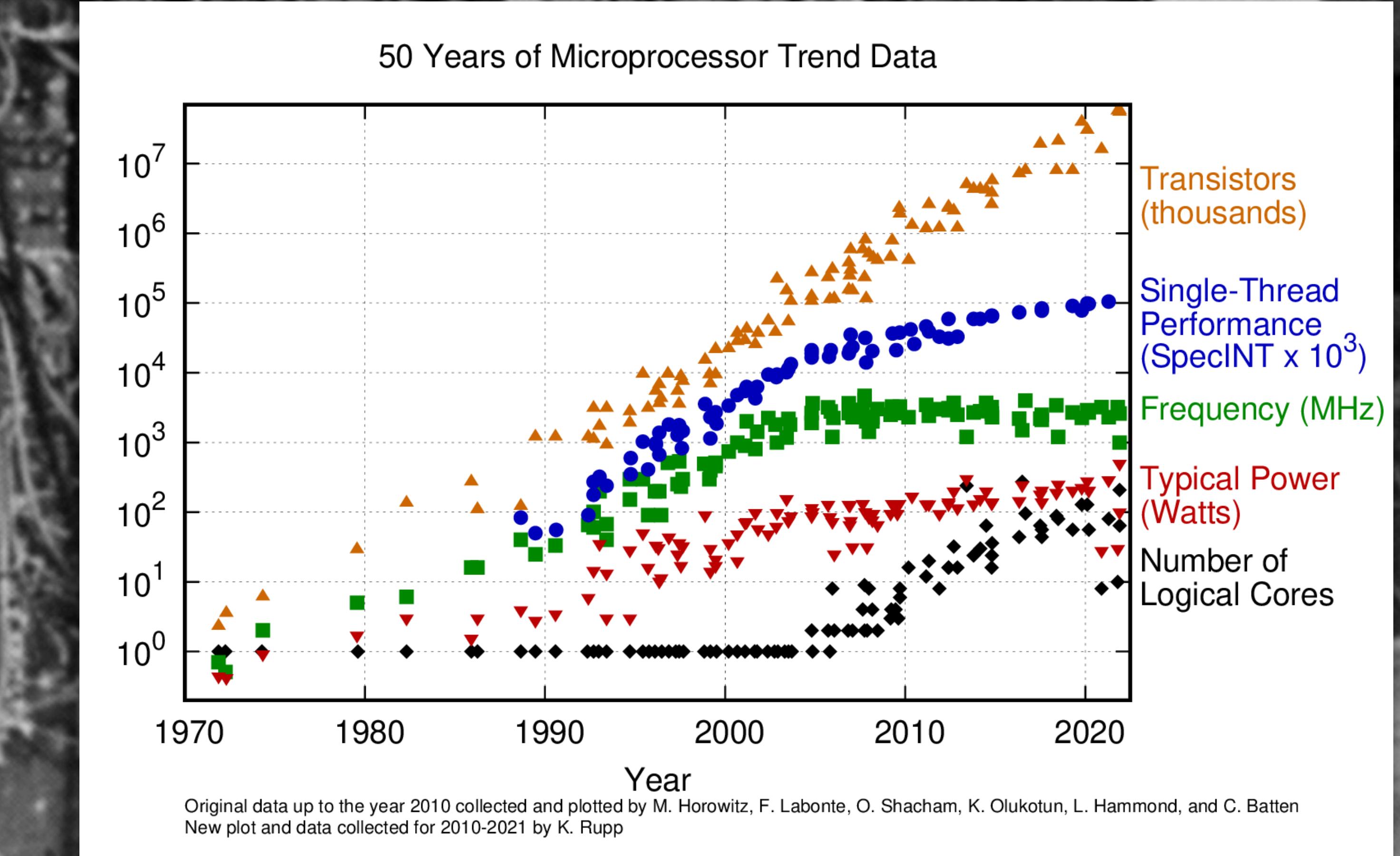


# Computational Science

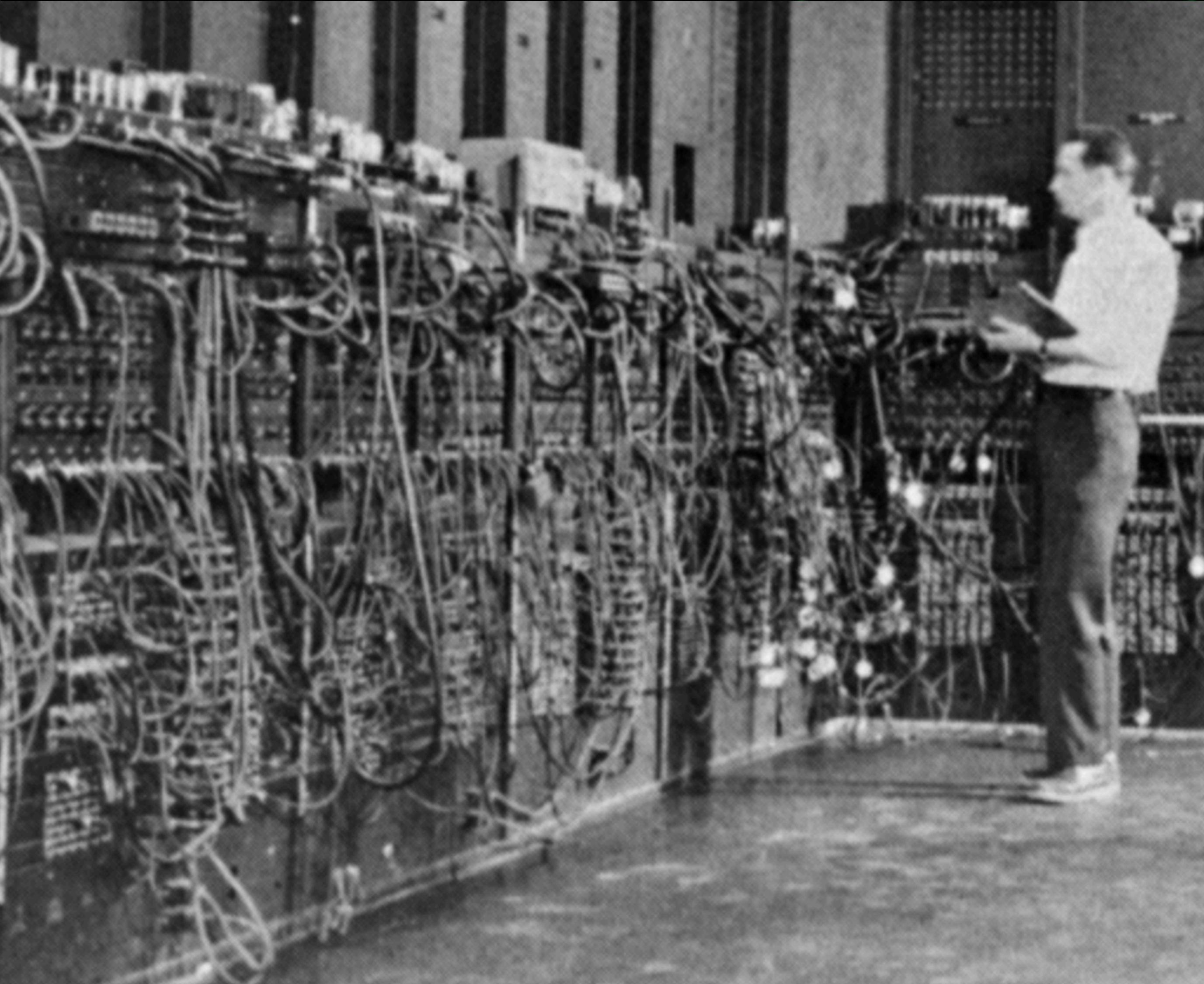


# Power of computing

We are entering “exascale” computing:  $10^{18}$  floating point operations per second



# Exponential growth of data



The Large Hadron Collider at CERN produced 40 zettabytes ( $10^{21}$ ) of data

Amazon Web Services store 500 exabytes ( $10^{18}$ ) of data

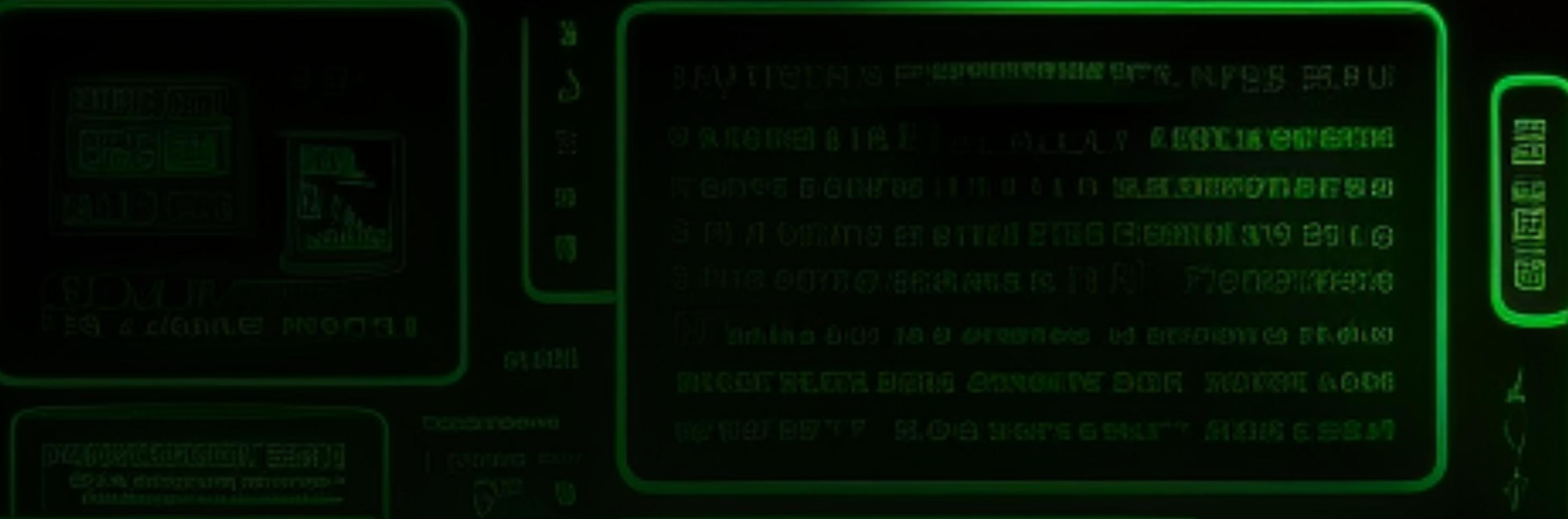
# Example: El Capitan



- 1.5 Exaflops
- Planned uses [1]:
  - *nuclear weapons*
  - *cancer research*
  - *additive manufacturing*
  - *climate*
  - *seismology*
  - *astrophysics*

[1] <https://techmonitor.ai/technology/worlds-most-powerful-supercomputer-cray-doe-nuclear>

# Could we simulate people instead of planets, atoms, and proteins?



# What is computational social science?



DATA SCIENCE COMPUTATIONAL SOCIAL SCIENCE

COMPUTER SCIENCE

SOCIAL SCIENCE

METHODS

# What is a social science?



DATA  
SOCIETE  
CULTURE  
ECONOMY  
INDIVIDUAL  
POLITICAL SYSTEM  
SOCIOLOGICAL THEORIES  
PSYCHOLOGICAL THEORIES  
POLITICAL THEORIES  
THEORIES  
DATA



# Social sciences

*“any branch of academic study or science that deals with human behaviour in its social and cultural aspects.”*

Economics

Anthropology

Sociology

Psychology

Political  
science

# History of quantitative social sciences



**1743 Condorcet:**  
“social mathematics”

**1749 Laplace:**  
Methods of physics  
to political science

**1796 Quetelet:**  
“social physics”

**1798 Comte:**  
“sociology”  
Social phenomena  
follows laws similar  
to natural phenomena

**1858 Durkheim:**  
- Father of modern social  
science  
- Social science methods  
equal to natural sciences

**1858 Simmel:**  
Social networks  
Emergence

# Conventional methods in social sciences

Surveys, questionnaires, observations: designing questions/experiments for targeted groups

- Labor intensive, small sample sizes, subjective answers, artificial settings

## Statistics as the quantitative method

- Significance testing = “how to deal with small sample size”
- Simplifying assumptions: e.g., linear regression

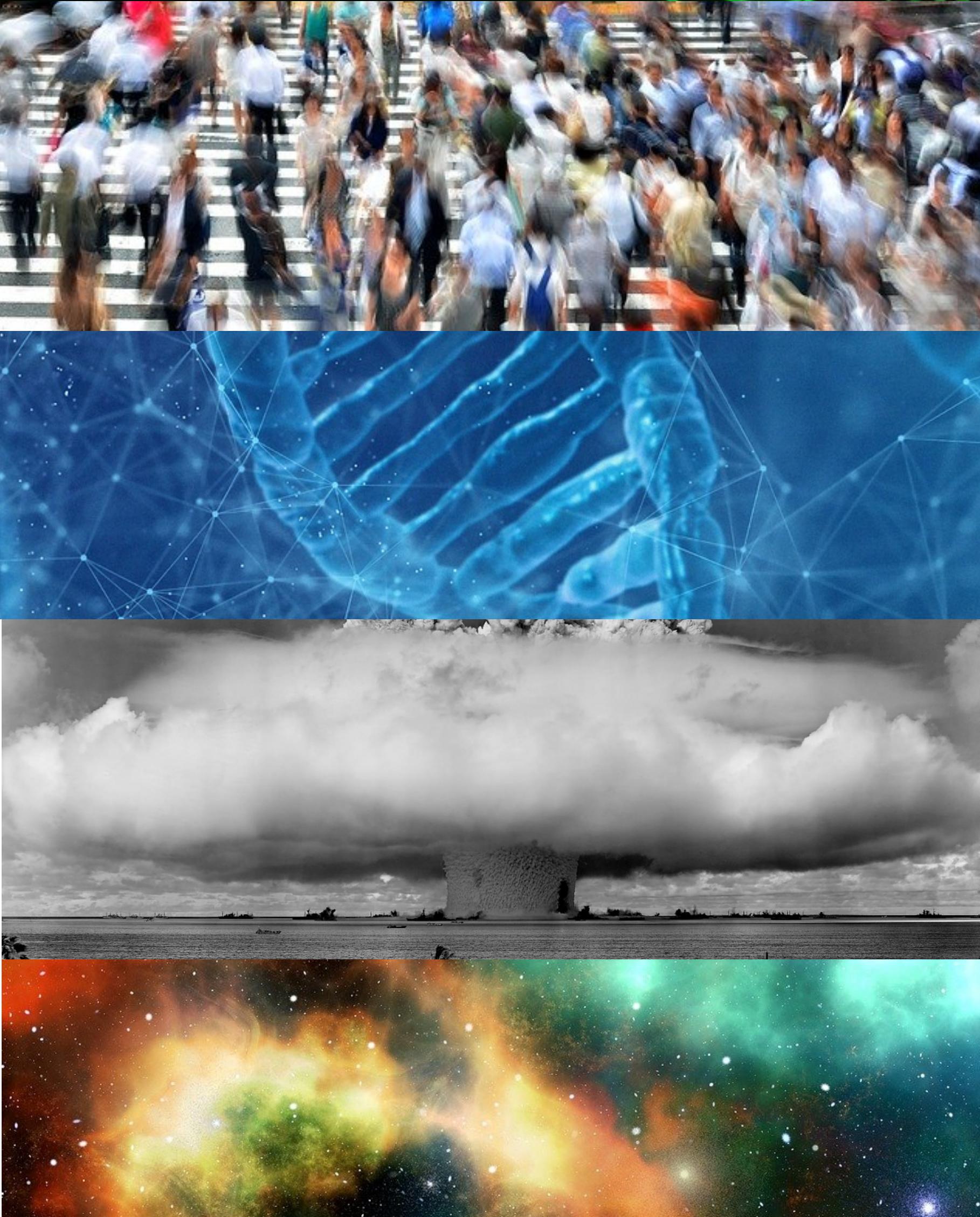
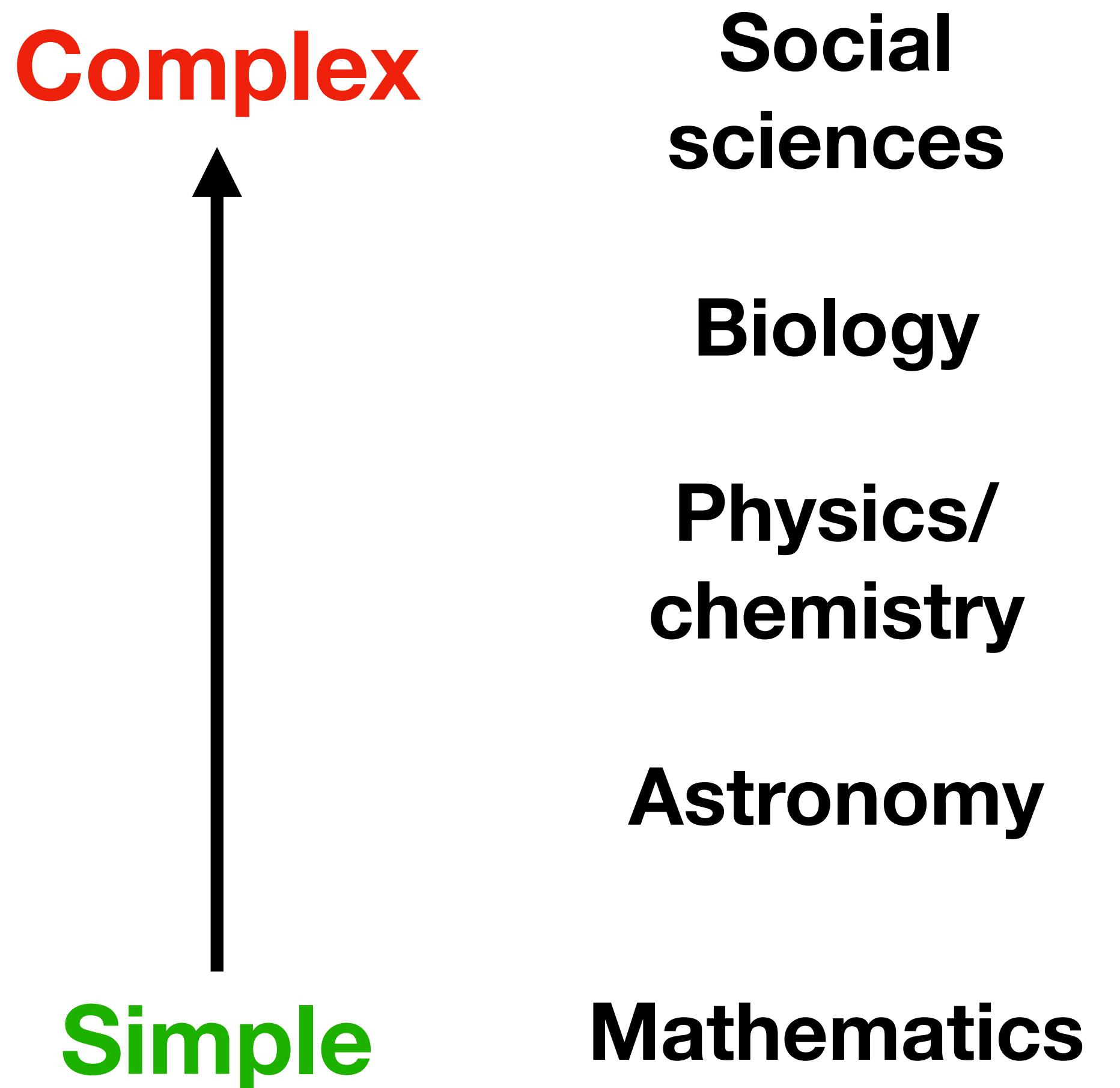
## Qualitative methods

- Our brains are excellent at finding patterns, making generalisations, etc
- Labour intensive, subjective

How to deal with large/structured data

How to deal with emergence and non-linearity?

# Hierarchy of sciences (Comte)



# Hierarchy of sciences (Comte)

**Complex**



**Social  
sciences**

**Biology**

**Physics/  
chemistry**

**Astronomy**

**Simple**

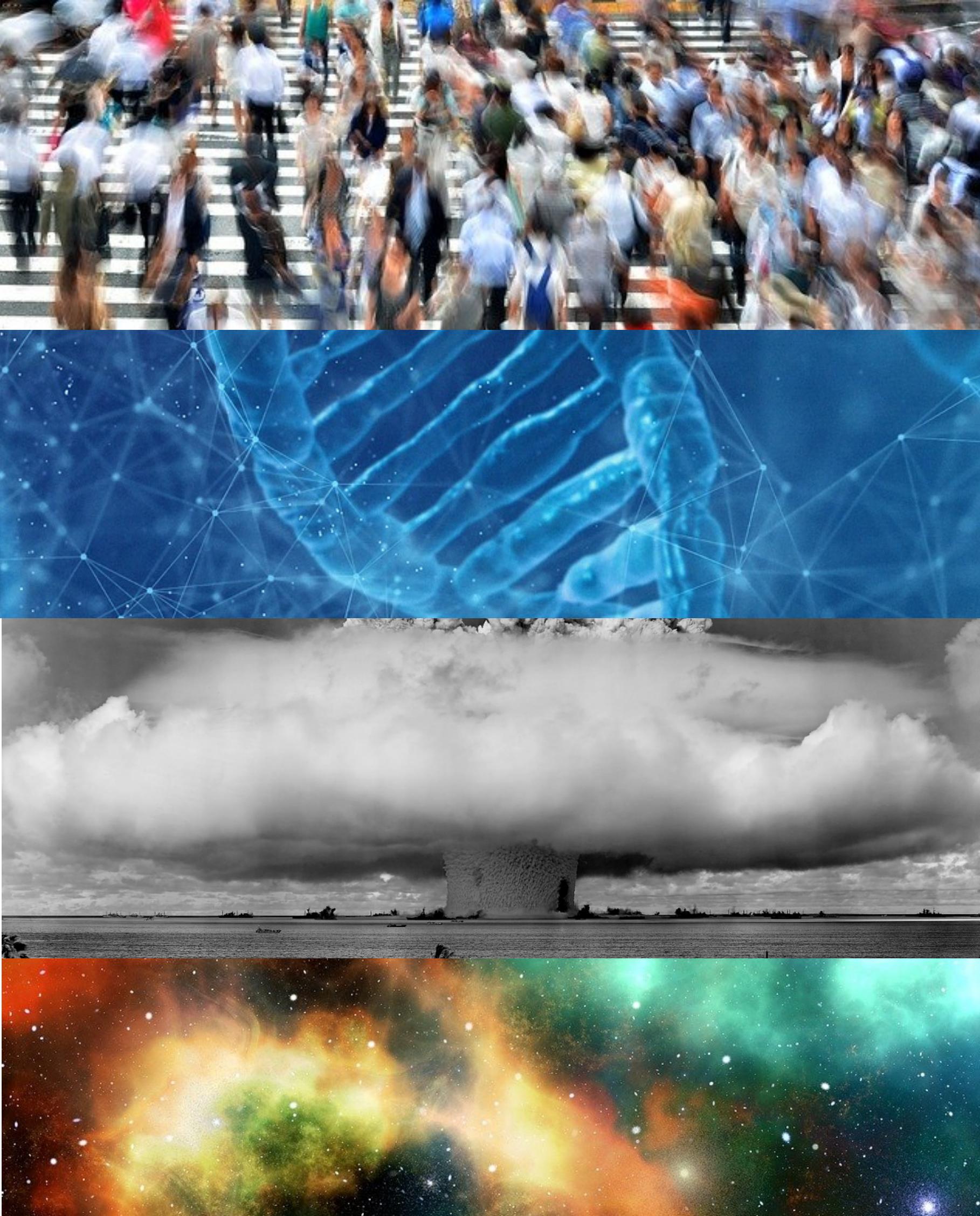
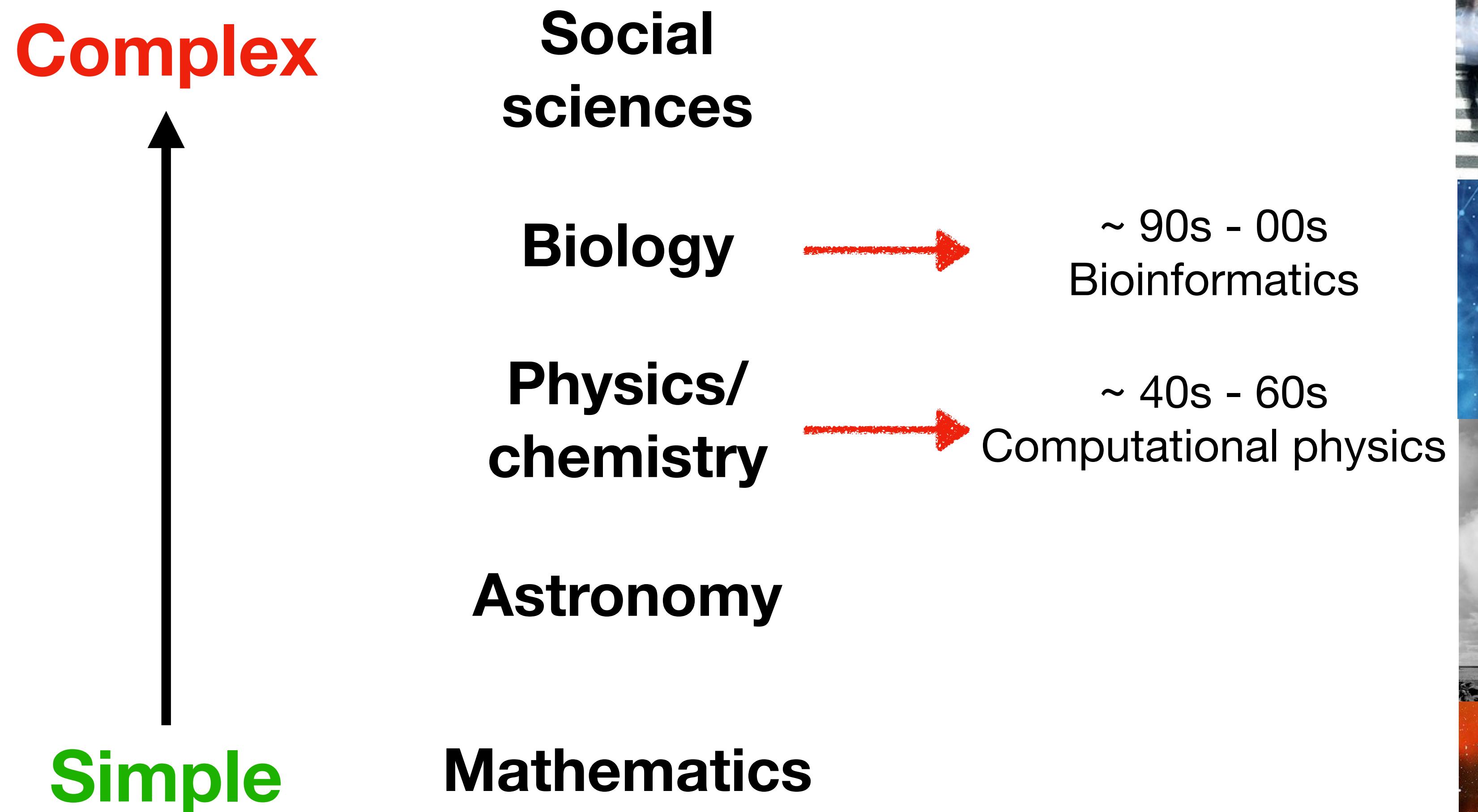
**Mathematics**



~ 40s - 60s  
Computational physics



# Hierarchy of sciences (Comte)



# Hierarchy of sciences (Comte)

**Complex**



**Social  
sciences**

**Biology**

**Physics/  
chemistry**

**Astronomy**

**Simple**

**Mathematics**



~ 10s - 20s  
Computational social  
science



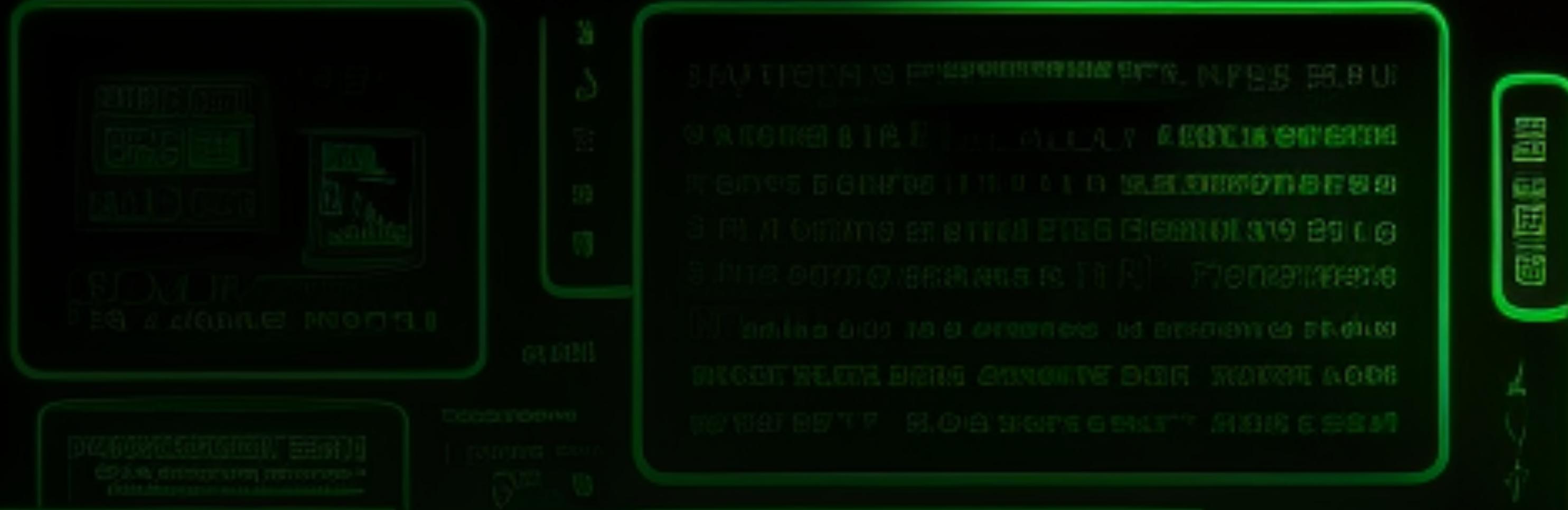
~ 90s - 00s  
Bioinformatics



~ 40s - 60s  
Computational physics



# What is computational social science?



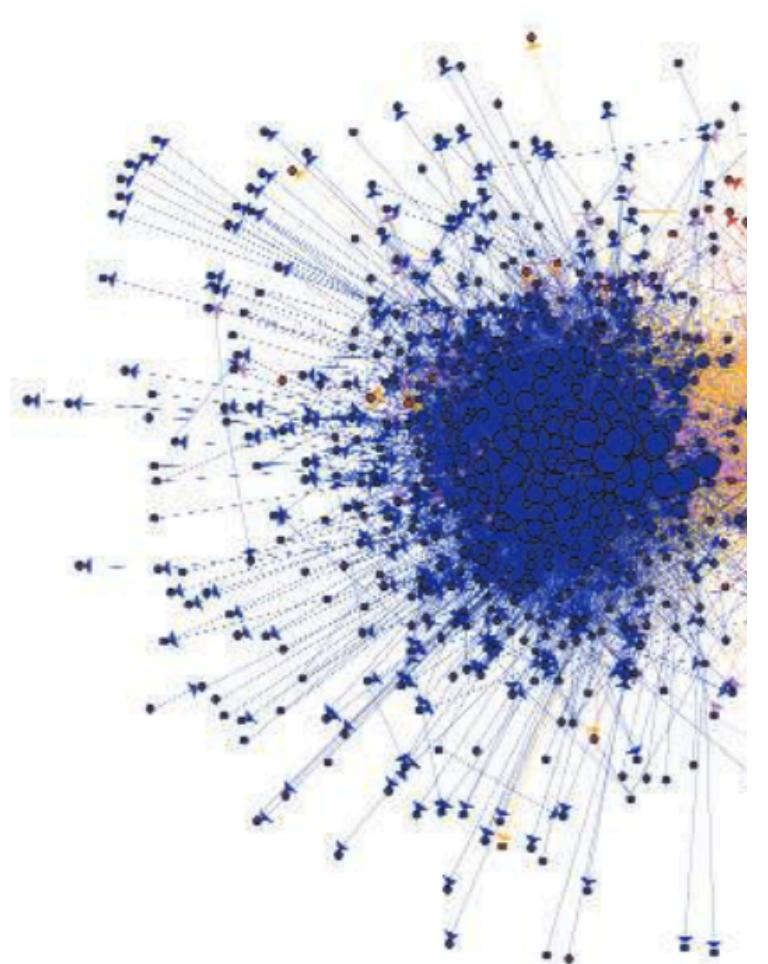
# Computational Social Science

David Lazer,<sup>1</sup> Alex Pentland,<sup>2</sup> Lada Adamic,<sup>3</sup> Sinan Aral,<sup>2,4</sup> Albert-László Barabási,<sup>5</sup> Devon Brewer,<sup>6</sup> Nicholas Christakis,<sup>1</sup> Noshir Contractor,<sup>7</sup> James Fowler,<sup>8</sup> Myron Gutmann,<sup>3</sup> Tony Jebara,<sup>9</sup> Gary King,<sup>1</sup> Michael Macy,<sup>10</sup> Deb Roy,<sup>2</sup> Marshall Van Alstyne<sup>2,11</sup>

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven “computational social science” has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in govern-

ment agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be



**Data from the blogosphere.** Shown is a link structure where red nodes indicate conservative blogs, and blue and purple ones from conservative to liberal. The size link to it. [Reproduced from (8) with permission from 1]

A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

*Annual Review of Sociology*

## Computational Social Science and Sociology

Achim Edelmann,<sup>1,2</sup> Tom Wolff,<sup>3</sup> Danielle Montagne,<sup>3</sup> and Christopher A. Bail<sup>3</sup>

<sup>1</sup>Institute of Sociology, University of Bern, 3012 Bern, Switzerland; email: achim.edelmann@soz.unibe.ch

<sup>2</sup>Department of Sociology, London School of Economics and Political Science, London WC2A 2AE, United Kingdom

<sup>3</sup>Department of Sociology, Duke University, Durham, North Carolina 27708, USA; email: christopher.bail@duke.edu

### THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

#### Regular Article

## Manifesto of computational social science

R. Conte<sup>1,a</sup>, N. Gilbert<sup>2</sup>, G. Bonelli<sup>1</sup>, C. Cioffi-Revilla<sup>3</sup>, G. Deffuant<sup>4</sup>, J. Kertesz<sup>5</sup>, V. Loreto<sup>6</sup>, S. Moat<sup>7</sup>, J.-P. Nadal<sup>8</sup>, A. Sanchez<sup>9</sup>, A. Nowak<sup>10</sup>, A. Flache<sup>11</sup>, M. San Miguel<sup>12</sup>, and D. Helbing<sup>13</sup>

<sup>1</sup> ISTC-CNR, Italy

<sup>2</sup> CRESS, University of Surrey, UK

<sup>3</sup> Center for Social Complexity, George Mason University, USA

<sup>4</sup> National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA), France

<sup>5</sup> Institute of Physics, Budapest University of Technology and Economics, Hungary

<sup>6</sup> Sapienza University of Rome, Italy

<sup>7</sup> University College London, UK

<sup>8</sup> CNRS, France

<sup>9</sup> GISC, Carlos III University of Madrid, Spain

<sup>10</sup> Center for Complex Systems, University of Warsaw, Poland

<sup>11</sup> ICS, University of Groningen, The Netherlands

<sup>12</sup> IFISC (CSIC-UIB), Campus Universitat Illes Balears, 07071 Palma de Mallorca, Spain

<sup>13</sup> ETH Zurich, Switzerland

Annu. Rev. Sociol. 2020. 46:61–81

First published as a Review in Advance on April 28, 2020

The *Annual Review of Sociology* is online at soc.annualreviews.org

<https://doi.org/10.1146/annurev-soc-121919-054621>

Copyright © 2020 by Annual Reviews.  
All rights reserved

### ANNUAL REVIEWS CONNECT

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

#### Keywords

computational social science, machine learning, network analysis, text analysis, demography, social psychology, economic sociology, political sociology, cultural sociology, sociology of knowledge

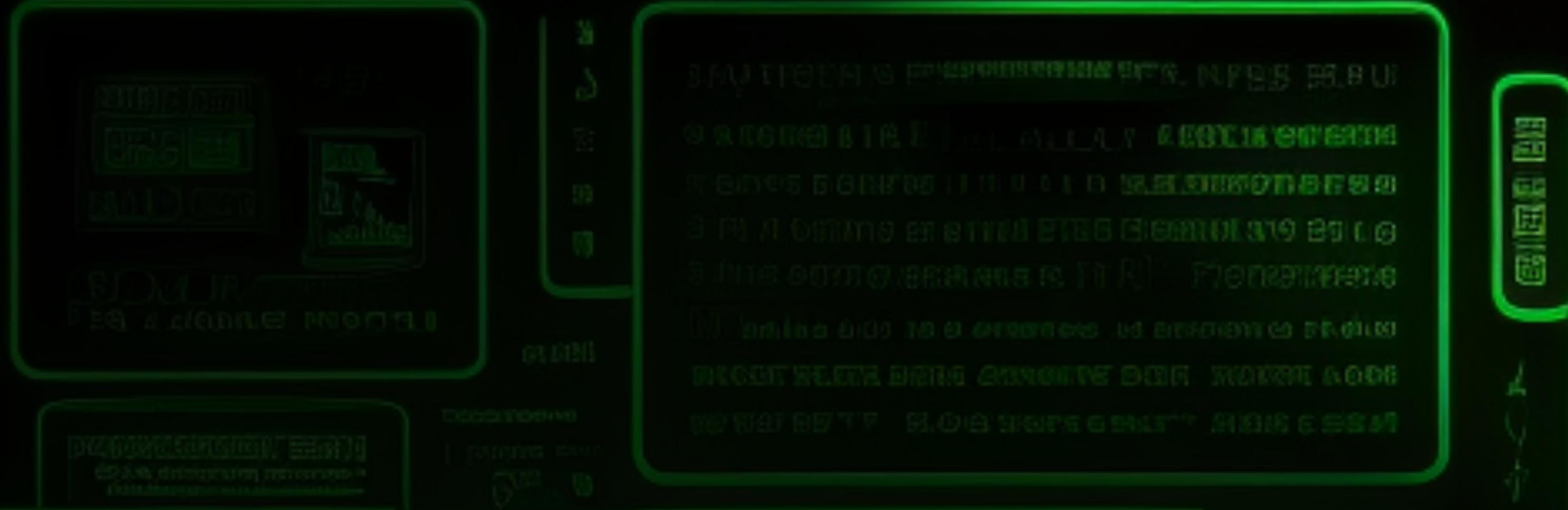
#### Abstract

The integration of social science with computer science and engineering fields has produced a new area of study: computational social science. This field applies computational methods to novel sources of digital data such as social media, administrative records, and historical archives to develop theories of human behavior. We review the evolution of this field within sociology via bibliometric analysis and in-depth analysis of the following subfields where this new work is appearing most rapidly: (a) social network analysis and group formation; (b) collective behavior and political sociology; (c) the sociology of knowledge; (d) cultural sociology, social psychology, and emotions; (e) the production of culture; (f) economic sociology and organizations; and (g) demography and population studies. Our review reveals that sociologists are not only at the center of cutting-edge research that addresses longstanding questions about human behavior but also developing new lines of inquiry about digital spaces as well. We conclude by discussing challenging new obstacles in the field, calling for increased attention to sociological theory, and identifying new areas where computational social science might be further integrated into mainstream sociology.

Received 1 August 2012 / Received in final form 9 October 2012  
Published online 5 December 2012

**Abstract.** The increasing integration of technology into our lives has created unprecedented volumes of data on society’s everyday behaviour. Such data opens up exciting new opportunities to work towards a quantitative understanding of our complex social systems, within the realms of a new discipline known as Computational Social Science. Against a background of financial crises, riots and international epidemics, the urgent need for a greater comprehension of the complexity of our interconnected global society and an ability to apply such insights in policy decisions is clear. This manifesto outlines the objectives of this new scientific direction, considering the challenges involved in it, and the extensive impact on science, technology and society that the success of this endeavour is likely to bring about.

# What is computational social science?



# Artificial societies & agent-based models

or

“how to build and play with your own artificial  
societies”

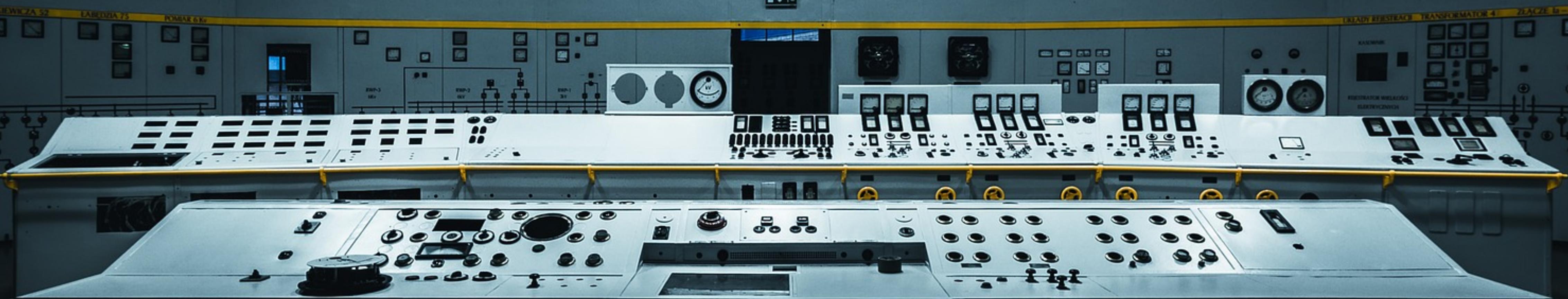
Data &  
digital traces





# Social networks

## Structure & dynamics



# Ethics, privacy, and legal issues

# Course structure

## Period IV

Week	Lecture	Exer. dl	Ext. dl	Topic
1	Feb 27	Mar 3	Mar 15	Introduction to CSS
2	Mar 6	Mar 10	Mar 22	Artificial societies & agent-based models
3	Mar 13	Mar 17	Mar 29	Data & digital traces
4	Mar 20	Mar 24	Apr 5	Counting things & analysing text
5	Mar 27	Mar 31	Apr 12	Social networks: structure
6	Apr 3	*	-	Introduction to the project

## Period V

Week	Lecture	Exercise dl	Ext. dl	Topic
7	Apr 24	May 5	May 10	Ethics, privacy, legal
-	-	-	-	WAPPU
8	May 8	May 12**	May 24	Agent-based models & emergence
9	May 15	May 19***	May 31	Social networks: dynamics
10	May 22	May 26***	June 7	Experiments & interventions at scale
11	May 29	-	-	Computing for social good

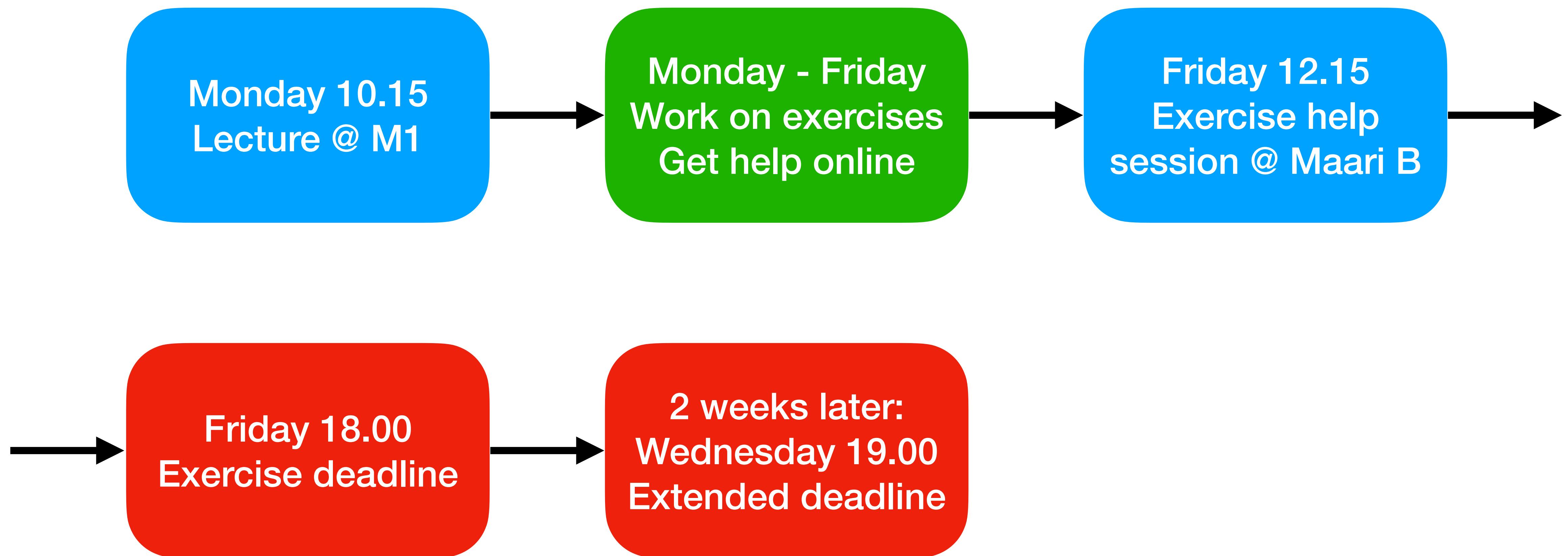
\*Project deadline: May 26

Project peer review: June 2

\*\*Bonus round

\*\*\*Only lecture questions

# Weekly schedule



# Exercises

- **Two types of exercises:**
  - Short multiple choice exercise related directly to the lecture
  - Programming + multiple choice
- **Exercises distributed through A+**
  - <https://plus.cs.aalto.fi/cs-e4730/2023spring/>
- **Exercises published on the day of the lecture, deadline following week Friday**
  - Extended deadlines = you get 80% of points

# Exercises

[A+ demo]

# Exercise design philosophy

- Programming: but focus not in programming techniques
- “Easy”: generous templates; focus on concepts, not technical details
- Frequent: new topic and problem set every week

# Communication channels



- 
- All messages
  - Recent topics
  - Private messages
  - Mentions
  - Starred messages
  - Drafts

STREAMS

▼	+
<b>1. exercise round</b> stream events	
2. exercise round	
3. exercise round	
4. exercise round	
5. exercise round	
exercise round: bonus	
exercise round: ethics, privacy, l...	
general	
project	

- Zulip chat, join via <https://css2023.zulip.aalto.fi/join/ezubrlv3quxurhink2heszxd/>
- If you have a questions that others might be wondering about, please ask in Zulip instead of email
- Monday lectures & Friday exercise sessions

# Project

- Idea: analyse a publicly available social dataset
- Include both fixed problems + open problems
- Peer grading (with checks from the course staff)
- Project is discussed more in lecture 6, at April 3rd
- Project deadline May 26th
- Project peer review deadline June 2nd

# Grading

- **70% exercises and lecture questionnaires**
  - 8.5 rounds with lecture questions, 60 points each: 510 points
  - $4 + 2 * 1/2$  rounds of exercise, 100 points each: 500 points
  - 1 bonus round of exercises, 100 bonus points
- **30% project**
  - Initial task: 100 points
  - Your own idea and reporting: 300 points
  - Peer review: 70 points

# Grading

- Total points: 1570 (no bonus or feedback)
- Point limits for grades: (subject to change, but not increased)

**1:** 780 points

**2:** 940 points

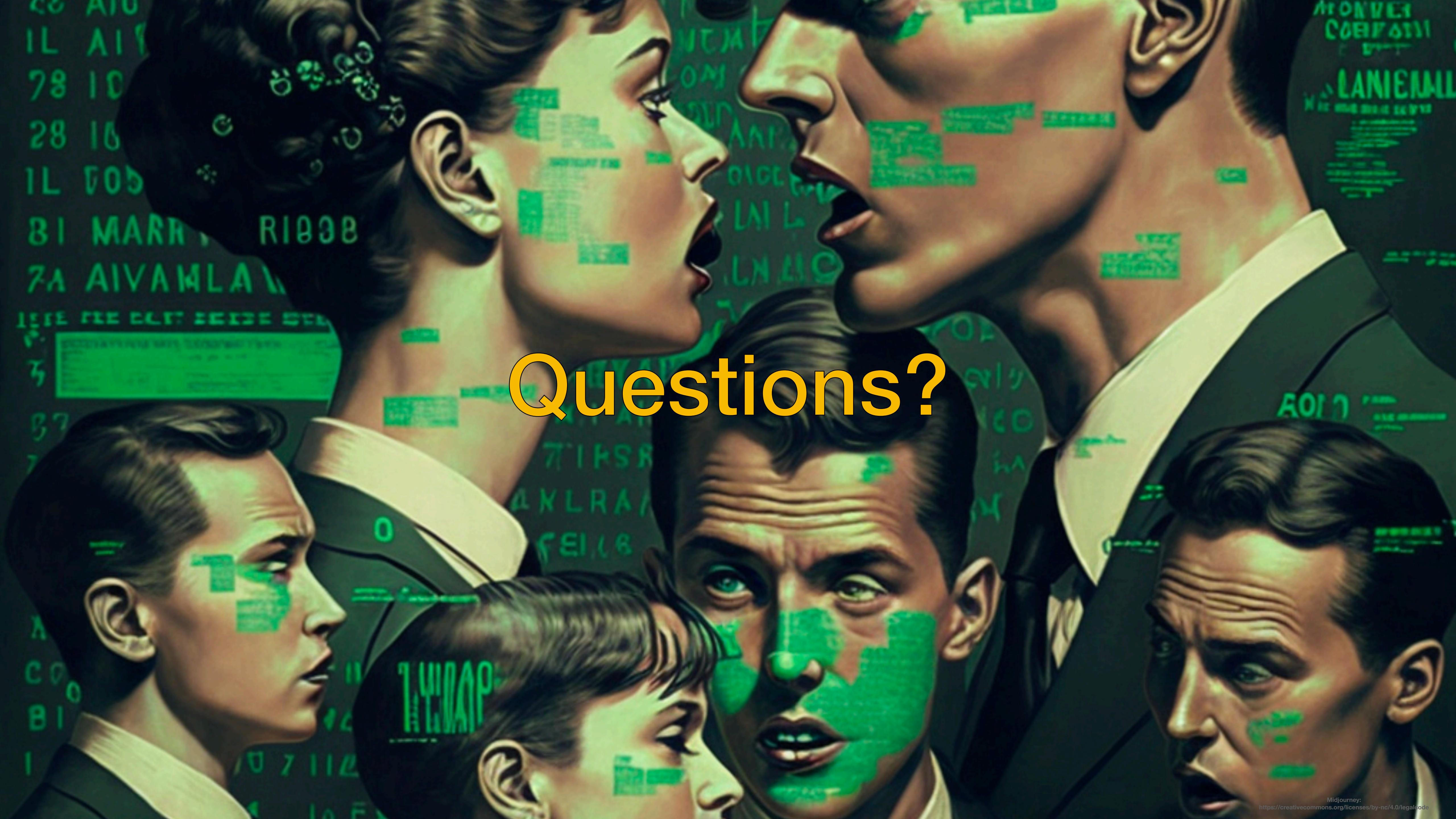
**3:** 1100 points

**4:** 1260 points

**5:** 1410 points

# Workload

- Lectures and lecture questions  $11 * 2h = 22h$
- Exercises:  $5 * 6h = 30h$
- Project: 30h
- Reflection: 51h
- Total: 133h



# Questions?